

## DOCUMENT INFORMATION EXTRACTING DEVICE

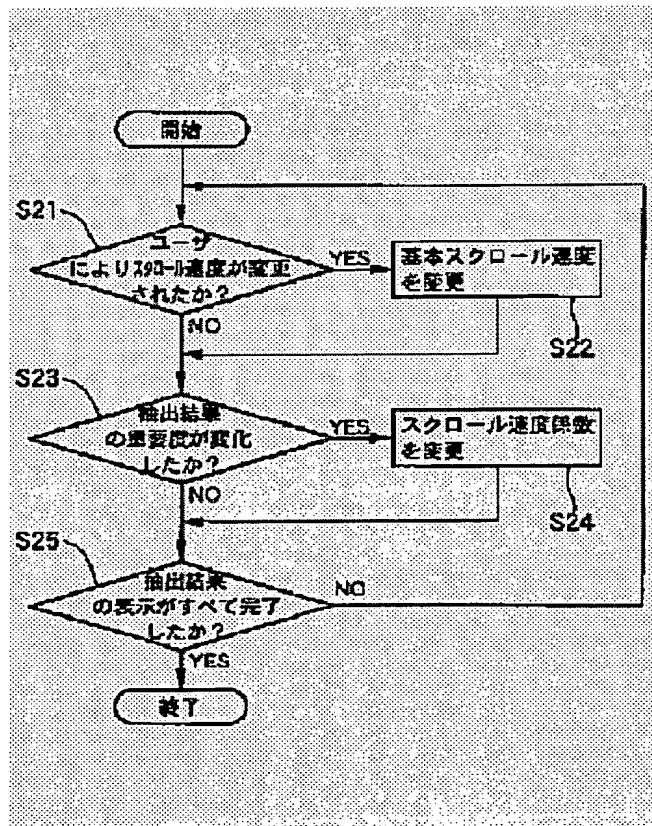
**Patent number:** JP2000099526  
**Publication date:** 2000-04-07  
**Inventor:** NISHIURA KAZUO  
**Applicant:** SHARP CORP  
**Classification:**  
 - international: G06F17/30; G06F3/00  
 - european:  
**Application number:** JP19980266105 19980921  
**Priority number(s):**

[Report a data error here](#)

### Abstract of JP2000099526

**PROBLEM TO BE SOLVED:** To provide a document information extracting device for controlling scrolling speed so that a user can easily read data at the time of displaying information extracted from document data on a limited display area.

**SOLUTION:** When the setting of scrolling speed is changed while an extraction result is displayed (step S21), basic scrolling speed is changed (step S22). When the criticality of the extraction result during display is changed (step S23), the factor of the scrolling speed coefficient to the basic scrolling speed is calculated based on the criticality (step S24) and the scrolling speed is transmitted as the product of the basic scrolling speed and the factor. When the display of all the extraction results completes (step S25), the control of the scrolling speed terminates.



Data supplied from the **esp@cenet** database - Patent Abstracts of Japan

(19)日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-99526

(P2000-99526A)

(43)公開日 平成12年4月7日 (2000.4.7)

(51)Int.Cl. <sup>7</sup> G 0 6 F 17/30 3/00	識別記号 6 5 6	F I C 0 6 F 15/403 3/00	テマコード <sup>8</sup> (参考) 3 8 0 A 5 B 0 7 5 6 5 6 D
---	---------------	-------------------------------	---

審査請求 未請求 請求項の数16 O L (全 10 頁)

(21)出願番号 特願平10-266105

(22)出願日 平成10年9月21日 (1998.9.21)

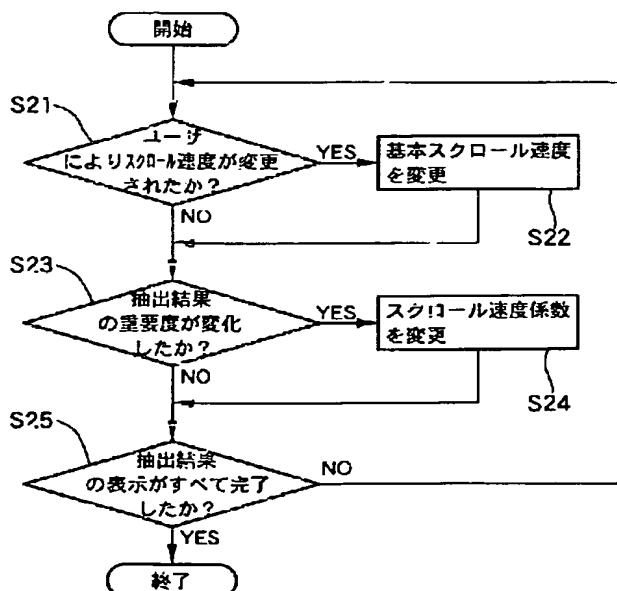
(71)出願人 000005049  
シャープ株式会社  
大阪府大阪市阿倍野区長池町22番22号  
(72)発明者 西浦 一夫  
大阪府大阪市阿倍野区長池町22番22号 シ  
ャープ株式会社内  
(74)代理人 100091096  
弁理士 平木 祐輔  
F ターム(参考) 5B075 MM02 MM11 PQ02 PQ22 PQ66  
PQ75

(54)【発明の名称】 文書情報抽出装置

(57)【要約】

【課題】 文書データから抽出された情報を限られた表示領域に表示する際に、ユーザにとって読み取りやすいスクロール速度となるように制御する文書情報抽出装置を提供すること。

【解決手段】 抽出結果の表示中、スクロール速度の設定の変更が行われた場合 (ステップS21) には、基本スクロール速度を変更する (ステップS22)。表示中の抽出結果の重要度が変化した場合 (ステップS23) には、重要度をもとに基本スクロール速度に対するスクロール速度係数を計算し (ステップS24)、スクロール速度を基本スクロール速度とスクロール速度係数の積として送出する。抽出結果の表示がすべて完了した時 (ステップS25) には、スクロール速度の制御を終了する。



【特許請求の範囲】

【請求項1】 文書データから特定情報を抽出する文書情報抽出手段と、該文書情報抽出手段によって抽出された特定情報（以下「抽出情報」という）を表示する表示手段とを備える文書情報抽出装置において、前記表示手段に前記抽出情報をスクロール表示させる表示制御手段と、該スクロール速度を変化させるスクロール速度制御手段とを備えることを特徴とする文書情報抽出装置。

【請求項2】 前記文書情報抽出手段は、重要度が所定値よりも高い文を抽出するものであることを特徴とする請求項1記載の文書情報抽出装置。

【請求項3】 前記文書情報抽出手段は、固有名詞を含む文を抽出するものであることを特徴とする請求項1記載の文書情報抽出装置。

【請求項4】 前記文書情報抽出手段は、日付表現を含む文を抽出するものであることを特徴とする請求項1記載の文書情報抽出装置。

【請求項5】 ユーザがスクロール速度を設定するスクロール速度ユーザ設定手段を備えることを特徴とする請求項1記載の文書情報抽出装置。

【請求項6】 前記表示手段が表示している抽出情報に応じてスクロール速度を制御するスクロール速度自動調節手段を備えることを特徴とする請求項1記載の文書情報抽出装置。

【請求項7】 前記表示手段が表示している抽出情報に応じて表示属性を制御する表示属性自動調節手段を備えることを特徴とする請求項1記載の文書情報抽出装置。

【請求項8】 前記文書データ又は抽出情報の言語の種類を判別する言語判別手段と、判別された言語が特定の言語以外である場合にその特定の言語に翻訳する翻訳処理手段とを備えることを特徴とする請求項1記載の文書情報抽出装置。

【請求項9】 情報抽出の対象とする複数の文書データを管理する情報抽出対象文書管理手段と、複数の抽出情報を1つにまとめる抽出情報結合手段とを備えることを特徴とする請求項1記載の文書情報抽出装置。

【請求項10】 前記複数の文書データの情報抽出の順序を決定する処理順序決定手段を備えることを特徴とする請求項9記載の文書情報抽出装置。

【請求項11】 前記複数の文書データの抽出情報の表示順序を決定する表示順序決定手段を備えることを特徴とする請求項9記載の文書情報抽出装置。

【請求項12】 情報抽出対象である文書データとその抽出情報との対応を管理する抽出結果対応管理手段と、ユーザからの入力指示を受けて、その時点で表示手段が表示している抽出情報に対し、前記抽出結果対応管理手段が管理している対応する元の文書データを表示手段に表示させる元文書データ表示指示手段とを備えることを特徴とする請求項1記載の文書情報抽出装置。

【請求項13】 所定の条件を満たしていない文書デー

タを抽出対象から除外する文書データフィルタリング手段を備えることを特徴とする請求項1記載の文書情報抽出装置。

【請求項14】 通信手段を介して文書データを取得するネットワーク文書取得手段を備えることを特徴とする請求項1記載の文書情報抽出装置。

【請求項15】 前記ネットワーク文書取得手段により取得した文書データがハイパーテキスト形式である場合、その文書データのリンク情報を解析してリンク先の文書データを取得するハイパーテキスト自動探索手段を備えることを特徴とする請求項14記載の文書情報抽出装置。

【請求項16】 コンピュータを請求項1乃至15いずれかに記載の文書情報抽出装置として機能させるためのプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、1つの、あるいは複数の文書データから特定の情報、例えば要約情報を抽出してユーザに提示する文書情報抽出装置において、特にユーザインターフェースに着目した文書情報抽出装置である。

【0002】

【従来の技術】従来、電子化された文書データから情報を抽出する技術として、最も単純な例の一つにK W I Cを挙げることができる。K W I Cとは key word in contextの略で、ある特定の単語、あるいは単語に限らず任意の文字列に関して、その文字列を含む前後の一定長の文字列を抽出し、表示を行う技術である。図7はK W I Cの実行結果例であるが、この図では文書データ中に出現する指定単語「文書データ」を中心としてその前後の文章の表示が行われている。K W I Cは特定の単語に関して、その単語が使用される前後の文脈の把握が容易であるため、情報検索装置におけるキーワード検索機能などにおいて非常に頻繁に利用される技術である。

【0003】文書データから情報抽出を行うもう少し複雑な例としては、文書中の重要文抽出技術や文書要約技術を挙げることができる。重要文抽出技術とは文字どおり、文書中で重要度が高いと思われる文を抽出し、表示を行う技術である。この技術は文書データ、特にサイズの大きい文書データに対して、重要度が高いと判定された文のみを抜き出して表示を行うことにより、装置の使用者がより少ない労力で文書データの概略把握を行えるような手段を提供することを目的とするものである。文の重要度の判定には、自然言語処理と呼ばれる技術に属する構文解析技術や意味解析技術が使用されることが多い。

【0004】文書要約技術は、単純に文書データから抽出した情報をそのまま表示するのではなく、抽出した情

報をさらに加工し、再構成してから表示する点に特徴がある。例えば文書データから重要と思われる単語や文、あるいは段落といった情報の抽出を行うのであるが、抽出した情報をそのまま並べて表示するのではなく、人間にとて読みやすい文となるような処理を加えてから表示する。

【0005】このような抽出情報に対する加工、再構成の処理にも自然言語処理技術が頻繁に使用される。例えば抽出した文間への接続詞の挿入、文脈処理による主語や目的語の補完、文の態の変換などである。このように文書要約技術は、正確には情報抽出にとどまる技術ではなく、抽出した情報を更に加工、再構成する技術をも含んでいるが、一般にはこの両者をあわせて情報抽出と呼ぶことも多く、本発明においてもこのような意味で情報抽出という言葉を用いることにする。

【0006】重要文抽出技術や文書要約技術に関する実現手段としてはすでにいくつもの方法が提案されており、例えば言語処理学会第4回年次大会ワークショップなどにおいても効率的、高精度な方法の提案が行われている。重要文抽出技術については、商用のシステムにおいてもワードプロセッサの1機能等として使用されている。またこれらの他にも、固有名詞を含む文の抽出や、日付表現を含む文の抽出など、様々な形での情報抽出装置が提案、実現されている。

【0007】このような情報抽出技術は、近年のワードプロセッサやパーソナルコンピュータ、あるいはインターネットの普及によって益々増大する傾向にある大量、大容量の電子化文書データの中から、ユーザーにとって必要な情報をすばやく探し出すという情報検索の目的、あるいは文書の概略をユーザーがすばやく理解できるような手段を提供するという速読支援の目的で使用されるのがこれまで一般的である。

【0008】そのためこのような技術を備えた文書情報抽出装置においては、抽出結果をユーザーに提示する際の表示の形態として、抽出結果の一覧性に重点が置かれることが多い。すなわち、大量の文書データの中から何らかの情報を取り出す必要に迫られているユーザーを前提とし、ユーザーがその目的をより効率よく、より短時間で達成できるような手段を提供するための表示形態である。

【0009】このような表示形態は大量、大容量の文書データを取り扱う際の労力を減少させ、作業の能率を向上させたいという欲求から生まれた実利的な方法であるが、近年のインターネットの普及による電子メール、電子掲示板、あるいはWWW(World Wide Web)と呼ばれるハイパーテキスト形式での文書データの流通は一方で、大量、大容量の文書データの閲覧、編集、管理を好み、それ自体に娛樂性を感じるユーザーをも生み出している。

【0010】

【発明が解決しようとする課題】このようなユーザーにとっては情報抽出技術自体が有用であり、インターネット

上の検索エンジンなどのように情報抽出技術が搭載された装置が頻繁に使用されることに変わりはないが、一覧性を重視した従来の表示形態は必ずしもユーザーの嗜好を満たすものではなかった。

【0011】またこれとは別に、例えば現在のパーソナルコンピュータにおいて主流となっているような、複数のプログラムモジュールを同時に実行可能であるマルチタスクオペレーティングシステム上で実現された文書情報抽出装置について考えてみると、ユーザーは装置が備える情報抽出の機能を利用する一方で、装置が備える別の機能、例えば文書作成のためのワードプロセッサ機能や、あるいはゲームの機能などを同時に実行するといった、いわゆる「ながら作業」となるような利用状況が頻繁に発生する。

【0012】このような利用状況において一覧性を重視した従来の表示形態では、表示手段において相当部分の表示領域を占有して結果の表示を行う必要があり、ユーザーが別の作業のために使用できる表示領域が限られたものになってしまうという問題があった。本発明は、上記問題点に鑑み、文書データから抽出された情報を限られた表示領域に表示する際に、ユーザーにとって読み取りやすいスクロール速度となるように制御する文書情報抽出装置を提供することを目的とする。

【0013】

【課題を解決するための手段】本発明の文書情報抽出装置は、文書データから特定情報を抽出する文書情報抽出手段と、該文書情報抽出手段によって抽出された特定情報を表示する表示手段とを備えるものであって、前記表示手段に前記抽出情報をスクロール表示させる表示制御手段と、該スクロール速度を変化させるスクロール速度制御手段とを備えるものである。

【0014】また、前記文書情報抽出手段は、重要度が所定値よりも高い文を抽出するものであることで、文書データ中の重要文を抽出する文書情報抽出装置において、抽出結果を読みやすい速度でスクロール表示することができる。さらに、前記文書情報抽出手段は、固有名詞を含む文を抽出するものであることで、文書データ中の固有名詞を含む文を抽出する文書情報抽出装置において、抽出結果を読みやすい速度でスクロール表示することができる。また、前記文書情報抽出手段は、日付表現を含む文を抽出するものであることで、文書データ中の日付表現を含む文を抽出する文書情報抽出装置において、抽出結果を読みやすい速度でスクロール表示することができる。

【0015】また、ユーザーがスクロール速度を設定するスクロール速度ユーザ設定手段を備えることで、スクロール表示の速度をユーザ設定可能とすることができます。また、前記表示手段が表示している抽出情報に応じてスクロール速度を制御するスクロール速度自動調節手段を備えることで、スクロール速度を抽出内容に応じて自動

的に調節することができる。また、前記表示手段が表示している抽出情報に応じて表示属性を制御する表示属性自動調節手段を備えることで、文字色、文字サイズ、文字フォントなどの表示属性を抽出内容に応じて自動的に調節することができる。

【0016】また、前記文書データ又は抽出情報の言語の種類を判別する言語判別手段と、判別された言語が特定の言語以外である場合にその特定の言語に翻訳する翻訳処理手段とを備えることで、外国語で記述された文書データに対しても母国語で抽出結果を表示することができる。また、情報抽出の対象とする複数の文書データを管理する情報抽出対象文書管理手段と、複数の抽出情報を1つにまとめる抽出情報結合手段とを備えることで、複数の文書データに対しても一度に情報抽出処理を行うことができる。また、前記複数の文書データの情報抽出の順序を決定する処理順序決定手段を備えることで、複数の文書データに対して情報抽出処理を行う際に、有用な順番で各文書データの情報抽出処理を行うことができる。

【0017】また、前記複数の文書データの抽出情報の表示順序を決定する表示順序決定手段を備えることで、複数の文書データに対して情報抽出処理を行った際に、有用な順番で各文書データからの情報抽出結果を表示することができる。また、情報抽出対象である文書データとその抽出情報との対応を管理する抽出結果対応管理手段と、ユーザからの入力指示を受けて、その時点で表示手段が表示している抽出情報に対し、前記抽出結果対応管理手段が管理している対応する元の文書データを表示手段に表示させる元文書データ表示指示手段とを備えることで、ユーザが指示した時点で表示されている情報抽出結果に対応する元の文書データを表示することができる。

【0018】また、所定の条件を満たしていない文書データを抽出対象から除外する文書データフィルタリング手段を備えることで、特定の条件を満たす文書データのみを情報抽出処理の対象とすることができます。また、通信手段を介して文書データを取得するネットワーク文書取得手段を備えることで、ネットワーク上に存在する文書データに対しても情報抽出処理を行うことができる。

【0019】また、前記ネットワーク文書取得手段により取得した文書データがハイパーテキスト形式である場合、その文書データのリンク情報を解析してリンク先の文書データを取得するハイパーテキスト自動探索手段を備えることで、ネットワーク上に存在するハイパーテキスト形式の文書データに関連する文書データについて一括して情報抽出処理を行うことができる。また、本発明はコンピュータを上記の文書情報抽出装置として機能させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体である。

【0020】

【発明の実施の形態】以下添付図面を参照しながら本発明の好適な実施の形態について詳細に説明する。図1は本発明の一実施の形態による表示の一例であり、時間経過とともに徐々に表示がスクロールされる様子を表している。表示領域に表示しきれない分量の情報を表示するときに、先頭部分から徐々に表示を行って、表示領域がいっぱいになった時点で先頭部分の情報の表示を消去し、消去した部分を詰めるようにして現在の表示位置を移動、再表示させ、それによって新しく利用可能となった表示領域に続きの情報を表示するという手順を繰り返して順に全情報を出力する手法がスクロール表示であり、このようなスクロール表示は、表示領域の限られた表示装置を持つおおかたの情報処理装置一般で非常に頻繁に用いられる手法である。表示を書き換えて続きの情報を表示するタイミングとしては、表示すべき情報が装置によって生成・用意された時点、あるいはユーザが何らかの指示手段によって指示を行った時点で行われるのが通常である。

【0021】図2は、本発明の一実施の形態の構成を示すブロック図である。記憶装置1は、磁気ディスク等の磁気記憶装置、CD-ROM等の光記憶装置、RAM等の半導体記憶装置など、一般の計算機において使用される記憶装置1でよく、文書データはこの記憶装置1のファイルシステム上に格納される。

【0022】文書情報抽出手段8は、文書データから何らかの情報を抽出するための手段であり、例えば文書データから特定の情報を抽出するためのアルゴリズムを記述したソフトウェア、およびそのソフトウェアを実行するCPUにより実現可能である。表示手段14にはCRTやLCDなど、リアルタイムで表示の更新が可能な一般的の表示装置を使用することができる。表示制御手段13は、表示手段14に対して表示情報を送出するための手段であり、例えば表示手段に対するドライバーソフトウェアとして実現可能である。このドライバーソフトウェアは、表示手段における表示内容を一定期間毎に更新できるように設計された通常のドライバーソフトウェアでよい。

【0023】スクロール速度制御手段15は、表示情報のスクロールが所定の速度となるよう表示制御手段13に対してタイミング信号を送出するための手段であり、例えば一定期間毎に割り込み信号を発生させるタイマー装置と、その割り込み信号をもとに表示装置14における表示更新のタイミングを制御するソフトウェアとして実現可能である。このソフトウェアは、割り込み信号をカウントし、一定カウント毎に表示制御手段13に対してタイミング信号を送るといったアルゴリズムのものでよい。

【0024】文書情報抽出手段8のアルゴリズムとしては様々なものが考えられるが、文書データ中から例えば重要度の高い文を抽出するアルゴリズムによって実現す

ることができる。特開平6-215049号公報において文書中の要約を抽出する手順が説明されており、例えばこのようなアルゴリズムを用いることで、実現することが可能である。文書情報抽出手段8を、文書データ中から固有名詞を含む文を抽出するアルゴリズムにより実現することもできる。

【0025】特公平7-7419号公報において、文書中の固有名詞をより高い精度で検出する手順が説明されており、例えばこのようなアルゴリズムを用いることで、実現することが可能である。文書情報抽出手段8を、文書データ中から日付表現を含む文を抽出するアルゴリズムにより実現することもできる。特開平5-158971号公報において、文書中の日付表現を柔軟に検出するための手順が説明されており、例えばこのようなアルゴリズムを用いることで、実現することができる。

【0026】スクロール速度ユーザ設定手段17は、ユーザの好みにスクロール速度を設定するための手段であり、これはキーボードなどユーザが設定を指示するための手段と、その設定内容を記憶するメモリーにより実現可能である。メモリーの内容をスクロール速度制御手段15が参照することにより、スクロール速度が調節される。

【0027】スクロール速度自動調節手段16は、文書情報抽出手段8による情報抽出時の情報をもとにしてスクロール速度を決定する。これは、例えば文書情報抽出手段8について考えると、情報抽出時に計算される各文の重要度をもとにし、重要度とスクロール速度を対応付けるアルゴリズムによりスクロール速度を決定するソフトウェアとして実現が可能である。

【0028】このアルゴリズムは、例えばスクロール速度を各文の重要度の逆数に比例する値として計算すれば、重要度の低い文については速い速度でスクロールが行われるため表示が一瞬で過ぎ去り、また重要度の高い文についてはゆっくりとした速度でスクロールが行われるためユーザはその情報をじっくりと読むことが可能となる。

【0029】表示属性自動調節手段18は、文書情報抽出手段による情報抽出時の情報をもとにして表示色や表示サイズ、表示フォントを変更したり、点滅表示させるなどの表示属性を決定し、その内容を表示制御手段13に送出する。これは、情報抽出時に計算される各文の重要度をもとにし、重要度の高い文のみ表示色を変更するといったアルゴリズムのソフトウェアとして実現が可能である。

【0030】言語判別手段9と翻訳処理手段10を備えることで、情報抽出結果を常に日本語で表示することが可能となる。言語判別手段9は、対象となる文または文書の文字コードや出現単語を調べるソフトウェアにより実現可能である。翻訳処理手段10は、構文解析処理部、格解析処理部、生成処理部などを備えた通常の翻訳

処理ソフトウェアなどとして実現可能である。

【0031】これら言語判別手段9、翻訳処理手段10は図2中に示されるように、文書情報抽出手段8と表示順序決定手段11(後述)の間に挿入することで、文書情報抽出結果に対して翻訳処理を行ってもよいし、あるいは処理順序決定手段7と文書情報抽出手段8の間に挿入することで、文書情報抽出処理を行う前に翻訳処理を行っても構わない。あるいは抽出情報結合手段12と表示制御手段13の間に挿入するといったことも可能である。言語判別手段9、翻訳処理手段10の挿入位置は自由度が高い。

【0032】情報抽出対象文書管理手段6は、情報抽出の対象とする複数の文書データを管理するための手段であり、それぞれの文書データへのファイルアクセスを制御するファイルポインタを管理するためのソフトウェアとして実現可能である。これにより、複数の文書に対して続けて情報抽出処理を行うことを可能とする。抽出情報結合手段12は、複数の文書データからの情報抽出結果を1つにまとめて表示制御手段13に送るための手段であるが、これは各文書データからの情報抽出結果を順番に続けてメモリーに書き込み、書き込まれたメモリーの内容を表示制御手段13へと受け渡すアルゴリズムを持ったソフトウェアとして実現可能である。これにより、一括して結果を表示することが可能になる。

【0033】処理順序決定手段7は、情報抽出対象文書管理手段6において管理されている文書データに対して、その処理順序を決定するための手段である。これは、例えば情報抽出対象文書管理手段6において管理されている文書データについて、文書データ作成日時をもとにソートを行うようなソフトウェアにより、作成日時の新しい順番で文書情報抽出手段8に文書データの内容を送出することで実現可能である。

【0034】表示順序決定手段11は、複数の文書データから文書情報抽出手段8によって抽出されたそれぞれの情報抽出結果に対して、その表示順序を決定する手段であるが、これは例えば文書情報抽出手段8において情報抽出時に中間に作成される抽出結果の重要度などをもとにし、重要度の高かった抽出結果から順に抽出情報結合手段12へと送信する、といったソフトウェアにより実現可能である。

【0035】抽出結果対応管理手段20は、複数文書データ処理時に、それぞれの文書データからの情報抽出結果と、その元文書データとの対応を保持するための手段である。これは、各文書データからの情報抽出結果と、その元文書のファイルポインタの対応関係をメモリー上に記憶させることで実現可能である。元文書データ表示指示手段19は、キーボードやマウスなど、一般的な入力装置により実現される。

【0036】元文書データ表示指示手段19よりユーザからの指示を受けたときには、抽出結果対応管理手段2

0が管理する対応情報をもとに現在表示装置に表示中の情報抽出結果に対応する元文書データのファイルポインタを取得し、この文書データの内容を表示制御手段13に対して送出することで元文書データを閲覧することが可能となる。これによりユーザは複数の文書データの抽出情報をスクロール表示で次々と眺めながら、必要な情報が出現した時点で即座に元の文書データを閲覧することが可能となる。

【0037】文書データフィルタリング手段5は、記憶装置1中に格納されている文書データのうち特定条件に従うものだけを情報抽出の対象とするための手段である。これは、例えば記憶装置1中に格納されている各文書データに対して文書データ作成日時を取得するソフトウェアにより、特定期間に作成された文書データのみを情報抽出対象文書管理手段6での管理対象とすることを実現可能である。あるいは、文書データの内容を検索して特定のキーワードを含む文書データを検出するソフトウェアにより、キーワードを含む文書データのみを情報抽出の対象とすることも可能である。通信手段2は、ネットワーク上に接続された他の計算機と通信を行う手段であるが、これにはイーサネットなど通常のネットワーク環境において使用されるハードウェア、およびそのドライバソフトウェアでよい。

【0038】ネットワーク文書取得手段3は、通信手段2によりアクセス可能となったネットワーク上の文書データに対して、その内容を読み出すソフトウェアとして実現可能である。通信手段2、ネットワーク文書取得手段3を備えることにより、他の計算機の記憶装置中に格納されている文書データであってもネットワークを介して本発明の装置が備える記憶装置1中に存在する文書データと全く同様に情報抽出処理を行うことが可能となる。

【0039】ハイパーテキスト自動探索手段4は、ネットワーク文書取得手段3により取得された文書データがハイパーテキスト形式である文書データである場合に、この文書内のリンク情報を解析することで、ネットワーク文書取得手段3で更に次々と文書データを取得するための手段である。これは、例えばインターネットにおいて頻繁に用いられるHTML形式のハイパーテキスト文書であればHTMLタグを解析し、解析したリンク情報をネットワーク文書取得手段3へと受け渡すソフトウェアとして実現可能である。これにより、ネットワーク上の計算機の記憶装置中に散在する大量の文書データに対して一括で情報抽出処理を行うことが可能となる。

【0040】図3、図4、図5、図6は、以上に述べた図2の構成を持つ情報抽出装置全体において、例えば1ヶ月以内に作成された文書データから情報抽出処理を行い、重要度順に結果の表示を行うような場合の処理フロー図である。ステップS1では本発明が備える記憶装置1、あるいは本発明と接続されたネットワーク上の文書

データから1ヶ月以内に作成された文書データのみを選択して取得する。これは図2においては記憶装置1、通信手段2、ネットワーク文書取得手段3、ハイパーテキスト自動探索手段4、文書データフィルタリング手段5、情報抽出対象文書管理手段6により行われる。

【0041】ステップS2では、情報抽出処理を行う順序を決定するために、文書データの作成日時順にソートを行う。この処理は文書データを処理順序決定手段7により行われる。ステップS3では、各文書データから情報抽出処理を行い。それぞれの抽出結果に対して重要度を算出する。この処理は文書情報抽出手段8により行われる。ステップS4では、ステップS3で処理が行われた抽出結果に対して、これが英文であるときには日本語へと翻訳処理を行う。この処理は言語判別手段9、翻訳処理手段10により行われる。

【0042】ステップS5では、各文書データからの抽出結果をステップS3で計算された重要度をもとにソートする。この処理は表示順序決定手段11により行われる。ステップS6では抽出結果を順に出力する。この処理は抽出情報結合手段12、表示制御手段13、表示手段14により行われる。抽出結果の表示中にユーザによって元文書表示指示が行われた場合(ステップS11)には、抽出結果対応管理手段20によって表示中の抽出結果に対応する元文書データが表示制御手段13へと送出される(ステップS12)。

【0043】抽出結果の表示がすべて終了した時(ステップS13)には、ユーザからの指示の検出を終了する。抽出結果の表示中、スクロール速度ユーザ設定手段17によってスクロール速度の設定の変更が行われた場合(ステップS21)には、基本スクロール速度を変更し(ステップS22)、スクロール速度制御手段15でその速度に応じたタイミング信号を表示制御手段13に送出する。

【0044】表示中の抽出結果の重要度が変化した場合(ステップS23)には、スクロール速度自動調節手段16において重要度をもとに基本スクロール速度に対するスクロール速度係数を計算し(ステップS24)、スクロール速度を基本スクロール速度とスクロール速度係数の積としてスクロール速度制御手段15はタイミング信号を表示制御手段13に対して送出する。この際、例えば、非重要文から重要文にスクロールしていくとき、重要文が表示に占める割合に応じてスクロールを徐々に遅くしたり、重要文が表示に占める割合が所定値に達したときにスクロールを遅くしたりすることができる。

【0045】抽出結果の表示がすべて完了した時(ステップS25)には、スクロール速度の制御を終了する。また、表示中の抽出結果の重要度が変化した場合(ステップS31)には、表示属性自動調節手段18において重要度をもとに表示文字サイズ等の表示属性を変更する(ステップS32)。変更された表示属性情報は、表示

制御手段に送出される。抽出結果の表示がすべて完了した時（ステップS33）には、表示属性の制御を終了する。なお、本発明は上記実施の形態に限定されるものではない。

〔0046〕

【発明の効果】請求項1に係る発明により、文書情報抽出装置において、抽出結果を読みやすい速度でスクロール表示し、動きのある結果表示によってユーザの注意を引き付けやすい、魅力的な表示形態とすることが可能となる。また例えはある文書の要約結果を、限られた表示領域の中でもユーザにとって読み取りやすい自然な速度となるようスクロール速度を制御しながら表示を行うことによって、マルチタスクオペレーティングシステム上での「ながら」作業の中でも、別の作業を行いながら快適に対象文書のすばやい内容把握が可能となるような文書情報抽出装置を構成することも可能となる。

【0047】請求項2に係る発明により、文書データ中の重要文を抽出する文書情報抽出装置において、抽出結果を読みやすい速度でスクロール表示する、新しい表示形態を提供することができる。請求項3に係る発明により、文書データ中の固有名詞を含む文を抽出する文書情報抽出装置において、抽出結果を読みやすい速度でスクロール表示する、新しい表示形態を提供することができる。請求項4に係る発明により、文書データ中の日付表現を含む文を抽出する文書情報抽出装置において、抽出結果を読みやすい速度でスクロール表示する、新しい表示形態を提供することができる。

【0048】請求項5に係る発明により、スクロール表示の速度をユーザ設定可能とすることができる。請求項6に係る発明により、スクロール速度を抽出内容に応じて自動的に調節することができる。請求項7に係る発明により、文字色、文字サイズ、文字フォントなどの表示属性を抽出内容に応じて自動的に調節することができる。

【0049】請求項8に係る発明により、外国語で記述された文書データに対しても母国語で抽出結果を表示す

することができる。請求項9に係る発明により、複数の文書データに対しても一度に情報抽出処理を行うことができる。請求項10に係る発明により、複数の文書データに対して情報抽出処理を行う際に、有用な順番で各文書データの情報抽出処理を行うことができる。

【0050】請求項11に係る発明により、複数の文書データに対して情報抽出処理を行った際に、有用な順番で各文書データからの情報抽出結果を表示することができる。請求項12に係る発明により、ユーザが指示した時点で表示されている情報抽出結果に対応する元の文書データを表示することが可能となる。請求項13に係る発明により、特定の条件を満たす文書データのみを情報抽出処理の対象とすることができます。

【0051】請求項14に係る発明により、ネットワーク上に存在する文書データに対しても情報抽出処理を行うことができる。請求項15に係る発明により、ネットワーク上に存在するハイパーテキスト形式の文書データに関連する文書データについて一括して情報抽出処理を行うことができる。請求項16に係る発明により、上記の文書情報抽出装置を記録媒体として実現することができる。

### 【図面の簡単な説明】

【図1】本発明の一実施の形態による表示の一例を示す図である。

## 【図2】本発明の一実施の形態の構成を示すブロック図である

### 【図3】本発明における、複数の文書データから情報抽出する処理フロー図である

【図4】本発明における、表示中の抽出情報に対応する元文書を表示する処理フロー図である。

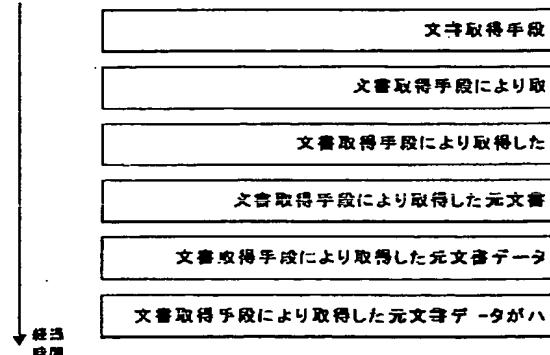
【図5】本発明における、スクロール速度を調節する処理フロー図である。

【図6】本発明における、表示属性を調節する処理フロー図である。

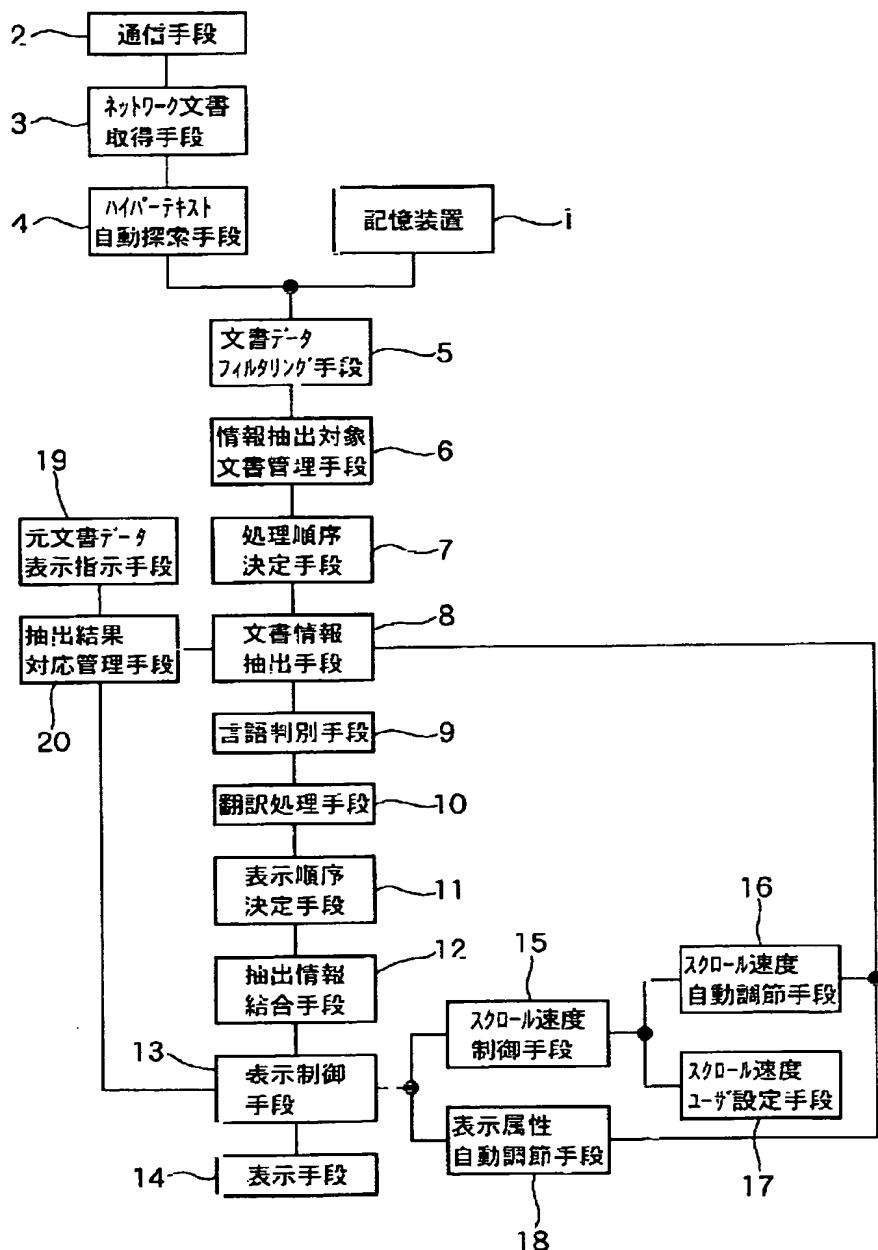
### 【図7】K W I C 形式による情報抽出結果の表示形態の一例を示す図である。

例を示す図である。

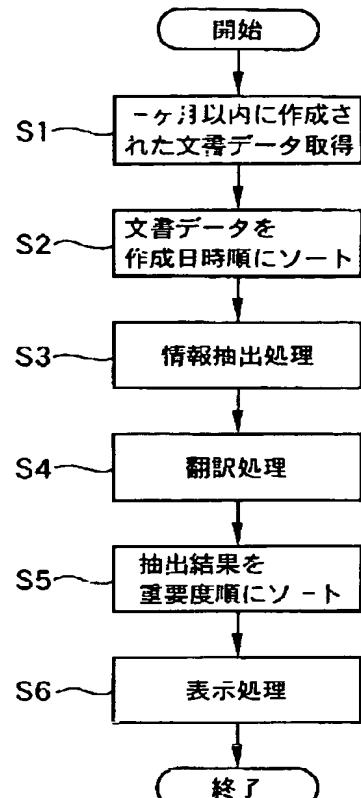
## 【図1】



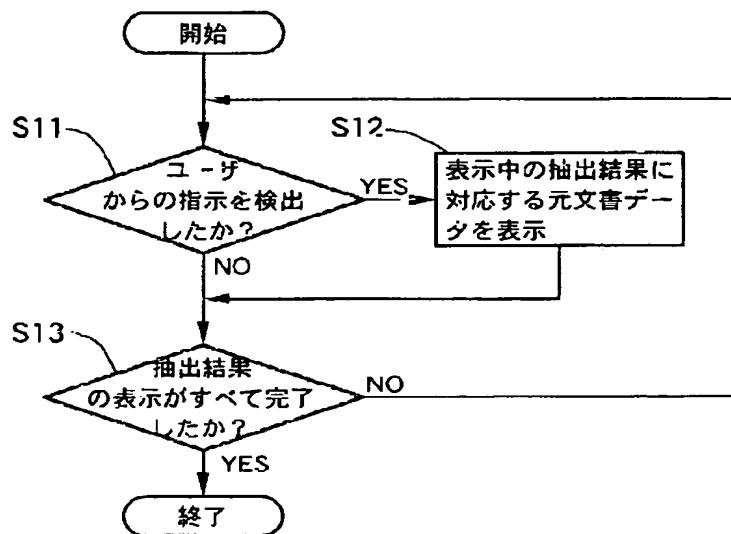
【図2】



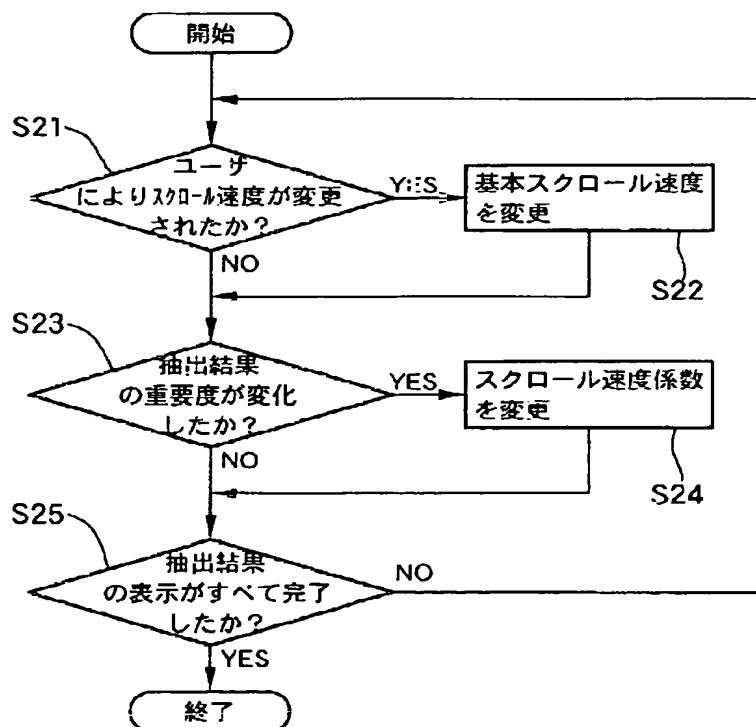
【図3】



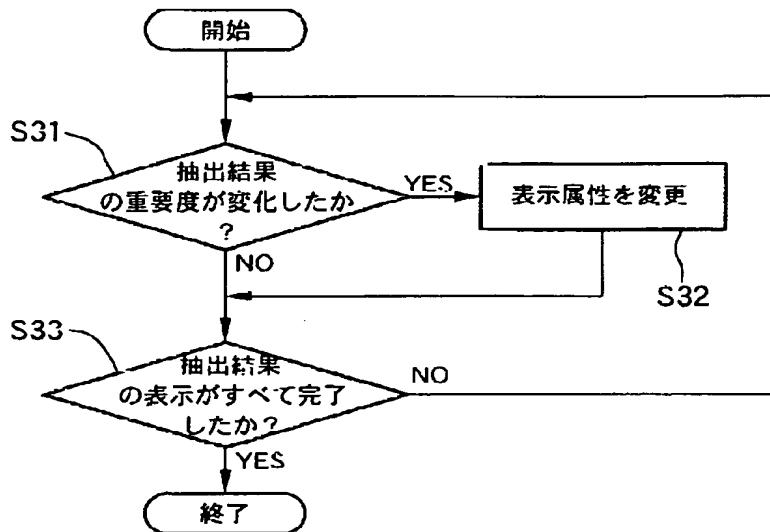
【図4】



【図5】



【図6】



【図7】

入力指示を受け取るための元文書データ表示指示手段を備え、ユーザから  
いる抽出情報に対応した元の文書データを表示手段に表示することを特  
徴とする文書情報抽出装置においては文書データフィルタリング手段を備え、記憶  
装置上に格納された文書データの内容を取得するネットワーク  
文書取得手段により取得した文書データがハイパーテキスト形式である  
場合から特定の情報、たとえば文書データの要約情報を抽出してユーザへ  
表示する結果例であるが、この図では文書データ中に出現する指定単語を中心と  
する技術である。この技術は文書データ、特にサイズの大きい文書データ  
データ、特にサイズの大きい文書データに対して、重要度が高いと判定さ